from all the others in the stratum that it belongs to. Since this difference is not accounted for in the equations used for estimation, resulting estimates are biased.

Instead of taking this approach, a state stratification should have been done to begin with. The whole CLLI sample size would then have been allocated according to a two-way stratification (by state and with the original strata). Just as with the single stratification, every stratum would need at least two CLLIs allocated to it. Since this was not done, the FCC will not be able to produce precise estimates at the state level, even though they made sure that every state was represented in the sample. The sample that the FCC has drawn is not representative by state.

Margin of Error

The margin of error is a measure of the precision of an estimator. It is usually the plus/minus part of a confidence interval of the form:

$$\tilde{X} \pm t \cdot s(\tilde{X}),$$

where $\tilde{X}$ is an estimator of some population quantity X, e.g., the total number of compliant line items in the CPR, or the total in-place cost associated with missing property. The quantity $s(\tilde{X})$ is the standard error of the estimator, and t is a multiplying factor that is determined by the distribution of the standardized quantity

$$\frac{\tilde{X} - X}{s(\tilde{X})},$$

and the confidence that one wants to have in the estimate. Typically, t is a percentile of the standard normal distribution or Student's t distribution. Most basic statistics books have tables for finding these values. Statistical software and spreadsheet programs can also be used.

Following the discussion in Cochran,[8] if $\tilde{X}$ has a normal distribution with mean X, and $s(\tilde{X})$ is well determined, then t comes from the standard normal distribution. These are two very important assumptions, and if they are not true, other types of error bounds need to be calculated using more advanced techniques.

The more well known situation occurs when $\tilde{X}$ has a normal distribution, but the sample size is not large enough for $s(\tilde{X})$ to be well determined. In this case, the degrees of freedom need to be considered, and Student's t distribution is used to find the multiplying factor.

In a situation where stratification has been used, one needs to consider the degrees of freedom provided by each stratum. The distribution of $s(\tilde{X})$ is in general too complicated to simply compute the degrees of freedom for each stratum in the usual way – taking the CLLI sample size

[8] See Cochran, W. G. (1967). *Sampling Techniques*, 3rd ed. Wiley, New York. Pp. 95 -96.

within the stratum minus one, i.e., (n_h - 1) – and then add them up across all strata. An approximate method of assigning an effective number of degrees of freedom to $s(\widetilde{X})$ has been worked out by Satterthwaite.[9]

Let $v(\widetilde{X})$ be the total variance of the estimator, and $v_h(\widetilde{X})$ the component of $v(\widetilde{X})$ from stratum h. Then the effective degrees of freedom can be approximated as

$$n_e = \frac{\{v(\widetilde{X})\}^2}{\sum_h \frac{\{v_h(\widetilde{X})\}^2}{n_h - 1}}.$$

The value of n_e always lies between the smallest of the values (n_h - 1) and their sum. For the audit described in the draft report, this value will lie between 1 and 21. Such values are too small for the normal distribution to be used.

Why is it that the central limit theorem does not apply when there is a relatively large total sample size of line items (1,152)? This is due to the two-stage design. The variance between CLLIs contributes much more towards total variance than the variances within CLLIs. Thus, the number of locations chosen plays an important role, and this number was chosen to be relatively small in the FCC sample design.

In the calculation section below, we show that the effective degrees of freedom is in the range 7 to 13, depending on the estimation method used and the scoring of the property records audited.

The draft report uses the multiplying factor 1.96, obtained from the standard normal distribution for a 95 percent two-sided confidence level. The table below shows the multiplying factor associated with different confidence levels from Student's t distribution with different degrees of freedom.

Degrees of Freedom	**One Sided Confidence Bounds**		**Two Sided Confidence bounds**	
$\mathbf{n_e}$	**95%**	**99%**	**95%**	**99%**
7	1.895	2.998	2.365	3.499
9	1.833	2.821	2.262	3.250
11	1.796	2.718	2.201	3.106
13	1.771	2.650	2.160	3.012

[9] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, pp. 110-114.

Notice that the multiplying factors for two-sided bounds at 95 percent confidence are larger than the value from the normal distribution – namely, 1.96. Thus, the reported margin of error for all estimates in the draft report needs to be increased.

The above analysis is only useful if the underlying distribution of the estimator is normally distributed. This is true probably for an estimator of the proportion of compliant records – although it should be confirmed. But it is not very likely that this will be true for estimators associated with dollar values. Very often the dollar values of a collection of items, such as the property records, are highly skewed, i.e., there is a relatively large number of small valued items, and a relatively small number of extremely large valued items. The distribution of an estimator based on a small sample size from such a population is usually skewed as well. Hence, it is not normal.

To learn more about the distribution of an estimator for dollar values, we conducted a simulation experiment that estimated the total in-place cost of the BellSouth hardwire line item population under study. This was done as follows:

1. Define a frame of CLLIs for which the total number of line items, and the total in-place cost is known. The frame should be divided into 11 strata just like the frame the FCC used for sampling. We were unable to create a sampling frame that produced a summary table exactly the same as that given in Appendix B page 7 of the draft report. For a summary of the frame we did use, and how it compares to the frame used by the FCC for the audits, see Table 1 at the end of this appendix. In our view the two are reasonably close.
2. Randomly select n_h out of the N_h CLLIs within each stratum, and record C_{hi}, the total in-place cost for selected CLLI i in stratum h.
3. Estimate the total in-place cost using

$$\hat{C}_R = \sum_{h=1}^{L} \frac{M_h}{M'_h} \sum_{i=1}^{n_h} C_{hi}\ .$$ [10]

[10] This estimator and the mean squared error equation that follows are equivalent, at the CLLI level, to the ones the FCC published in Appendix B of the draft audit report. See the next section for a full description of the estimator.

4. Estimate the mean squared error of $\hat{C}_R$ using

$$v(\hat{C}_R)=\sum_{h=1}^{L}\left(\frac{N_h^2(1-f_{1h})}{n_h}\cdot\frac{\sum_{i=1}^{n_h}M_{hi}^2\left(\bar{c}_{hi}-\hat{\bar{\bar{C}}}_{Rh}\right)^2}{n_h-1}\right), \text{ where}$$

$$\bar{c}_{hi}=\frac{1}{M_{hi}}C_{hi}$$

$$\hat{\bar{\bar{C}}}_{Rh}=\frac{1}{M'_h}\sum_{i=1}^{n_h}C_{hi}, \text{ and}$$

$$f_{1h}=\frac{n_h}{N_h}.$$

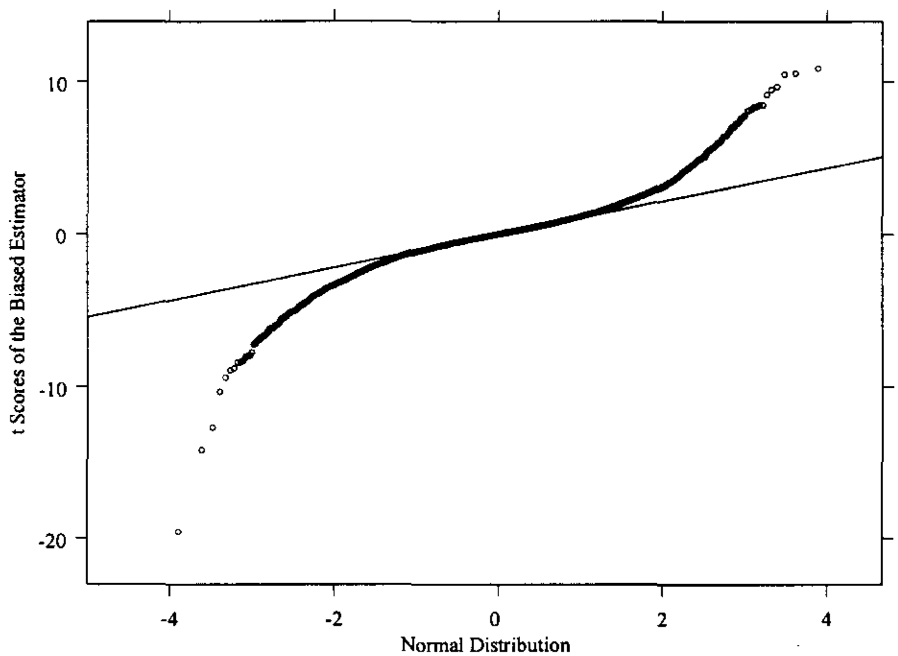
5. Calculate a t-score for the estimate, i.e., find the error in each estimate, $\hat{C}_R$ - C, where C is the known total in-place cost, and divide the error by $\sqrt{v(\hat{C}_R)}$.
6. Repeat steps 1 through 5 a large number of times. In our case we did 10,000 runs.

While this simulation does not perform an evaluation of the exact estimator the FCC used to estimate values the audit was interested in, it does provide information about how well the type of estimator that was used performs in estimating the in-place cost associated with non-locatable line items. This is because the simulation looks at estimates of a similar quantity, total in-place cost.

The simulation results give us an indication of how to proceed with determining a one-sided lower confidence bound by examining the distribution of the 10,000 realizations of the t-scores. We first compare the t-score distribution with a normal distribution via a normal q-q plot. This plot provides a powerful, visual comparison of the estimated quantiles of the t-scores with the same quantiles of a standard normal distribution. If the t-scores come from a normal (or nearly normal) distribution, then the resulting plot should look like a straight line. We follow Cleveland's[11] method of presentation where a reference line passing through upper and lower quartiles is "superposed" on the graph.

[11] Cleveland, W. S. (1993) *Visualizing Data.* Hobart Press, Summit, New Jersey.

Normal Q-Q Plot for t-Scores of $\hat{C}_R$



This plot tell us that both tails of the distribution are much heavier than that of a normal distribution – much like Student's t distribution.

To find multipliers for the root mean squared error so that we can obtain one-sided lower confidence bounds, we can use the 1 percent or 5 percent quantiles of the t-score distribution. These are presented below.

Item	1%	5%
t-score	-4.447	-2.434

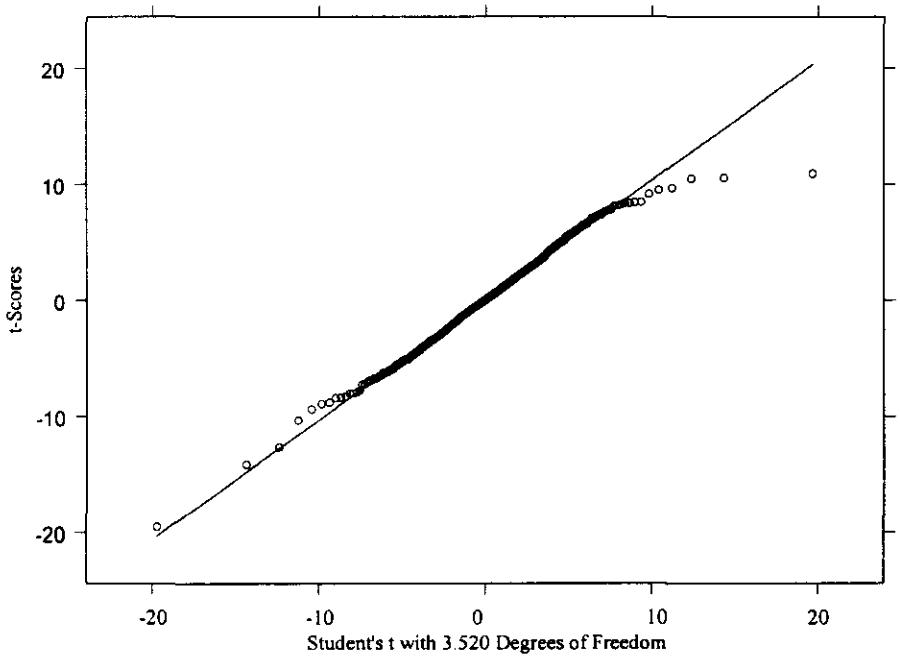
We can also use the results to answer the following questions.

1. Can Student's t distribution be used to find the multiplying factor for determining a lower confidence bound?
2. Is the Satterthwaite approximation for the effective degrees of freedom good?

To answer the first question, we proceeded by first identifying the degrees of freedom for a Student's t distribution that fit the t-scores. This was done by finding a least squares fit between the quantiles of the t-scores and a t distribution. We found that a t distribution with 3.520 degrees of freedom provides a least squares fit. To evaluate this fit, we compared the quantiles of the t-scores with those of a t distribution via the q-q plot shown below.

Notice, that the t distribution fits the t-scores fairly well in the middle of the distribution, and in the lower tail. The fit in the upper tail is poor though. Thus, we conclude that Student's t distribution may be adequate for finding the margin of error associated with lower bounds (which is what we need), but not for upper bounds.

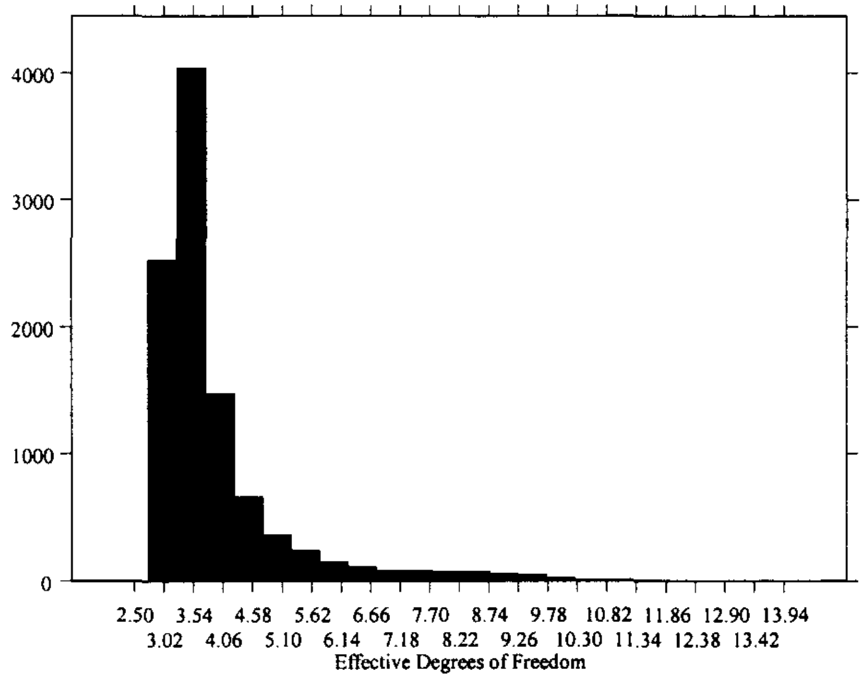
Student's t Q-Q Plot for t-Scores of $\hat{C}_R$



To determine how well the Satterthwaite approximation performs, we calculated the effective degrees of freedom for each of the 10,000 realizations of $\hat{C}_R$. A histogram of the distribution is shown below. The distribution is skewed, but most values are in the range from 3 to 4. The mean of the distribution is 3.917, and the standard deviation is 1.212.

The Satterthwaite effective degrees of freedom calculation appears to capture the t distribution found via the least squares fit – at least on average. There is a sizable chance (25 percent) that in a particular sample the effective degrees of freedom calculated could be too high (greater than 4).

Histogram of 10,000 Realizations of the Effective Degrees of Freedom



In light of the simulation results, we computed lower bounds for biased estimates of dollar values using the appropriate quantile of the t-score distribution given above, e.g., a t value of -4.447 for the 99 percent lower bound. We do not know if this analysis is compatible with proportion estimation.

More advanced techniques such as balance repeated replication, or the jackknife can also be used to determine error bounds in these more complex situations.[12] We will not go into these methods here, since we believe our point about the increase in the size of the margin of error has been made.

Once a correct approach is found for calculating error bounds, we have argued that a one-sided lower confidence bound should be used as the value assessed to be in error, e.g., the total in-place cost of non-locatable line items, or the proportion of non-compliant records. This is because only values smaller than the lower bound are, statistically speaking, significantly different from values above the lower bound.[13] As noted in the main summary, the IRS uses such a rule for its audit findings.[14]

[12] See Cochran, chapter 11, sections 18 - 20. See also, Wolter, K. M. (1985). *Introduction to Variance Estimation.* Springer-Verlag, New York.

[13] This concept comes from the statistical theory of hypothesis testing. Suppose, for example, that a standard of "materiality" was set for the audit (as should have been done). Suppose further that the level of materiality was set at 100 million dollars for in-place cost associated with non-locatable line items (a very small fraction of the hardwire total investment). Then one would test the null hypothesis that the true value is less than or equal to 100 million dollars versus the alternative that there is more than 100 million dollars associated with non-locatable line items. The deciding factor of this test is whether or not a one-sided lower confidence bound is below or above

Also, if one is going to take a conservative approach, the confidence level for this bound should be set at 99 percent. This practice attempts to take into account the uncertainty caused by various unquantifiable errors introduced into both the sampling and audit processes. In other words, as far as the FCC estimates of dollar values are concerned, use of a 99 percent lower bound of the proper confidence interval would be the prudent approach.

Sources of Bias that Affect the Estimates

Several forms of bias are present in the estimates supplied in the draft report. These include:

- the use of a statistically biased estimator,
- bias caused by substituting CLLIs and line items for undesirable ones that turned up in the sample, and
- biases induced by weaknesses in audit controls.

The effect of each of these biases varies in its degree of severity. The total effect may be significant; it certainly brings up legitimate concerns for the accuracy of the audit results. We address each in turn below.

Estimator Bias

The estimator used by the FCC is statistically biased. The FCC estimator can be useful in many situations, since it may have a smaller mean squared error than the standard unbiased estimator. The formula for this FCC estimator of a total population value is given by

$$\hat{Y}_R = \sum_{h=1}^{L} \frac{M_h}{M_h'} \sum_{i=1}^{n_h} \frac{M_{hi}}{36} \sum_{j=1}^{36} y_{hij} = \sum_{h=1}^{L} \frac{M_h}{M_h'} \sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi} = \sum_{h=1}^{L} \hat{Y}_{Rh}, \text{ where}$$

$$\hat{Y}_{Rh} = \frac{M_h}{M_h'} \sum_{i=1}^{n_h} M_{hi} \bar{y}_{hi}, \text{ and}$$

$$\bar{y}_{hi} = \frac{1}{36} \sum_{j=1}^{36} y_{hij}.$$

If y_{hij} is the in-place cost of an audited line item that was not located and zero otherwise, then $\hat{Y}_R$ is an estimator for the total in-place cost of non-locatable line items. On the other hand, if y_{hij} is one or zero depending on whether or not an audited line item is or is not compliant, then $\hat{Y}_R$ is an estimator for the total number of compliant line items in the population. If this is

whatever materiality standard is set. If it is below, then one could not statistically conclude that the true value is materially different from the value carried on the CPR.

[14] See footnote 1 of the summary report for which this appendix is an attachment.

divided by M_0 then we have an estimator for the proportion of compliant line items in the population.

In order to judge the precision of an estimator, statisticians usually look at the variance of the estimator, or its square root – the standard error of the estimator. For a biased estimate, the variance does not capture the precision of the estimator with respect to the true value of the population that is being estimated. The more appropriate measure is the mean squared error of the estimator, and its square root – referred to as the root mean squared error.

An approximate sample estimate for the mean-squared error for this estimator is given by

$$v(\hat{Y}_R)=\sum_{h=1}^{L}\left(\frac{N_h^2(1-f_{1h})}{n_h}\cdot\frac{\sum_{i=1}^{n_h}M_{hi}^2\left(\bar{y}_{hi}-\hat{\bar{\bar{Y}}}_{Rh}\right)^2}{n_h-1}+\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{M_{hi}^2(1-f_{2hi})s_{2hi}^2}{36}\right), \text{ where}$$

$$\hat{\bar{\bar{Y}}}_{Rh}=\frac{\hat{Y}_{Rh}}{M_h},$$

$$f_{1h}=\frac{n_h}{N_h},\quad f_{2hi}=\frac{36}{M_{hi}} \text{ and}$$

$$s_{2hi}^2=\frac{1}{35}\sum_{j=1}^{36}\left(y_{hij}-\bar{y}_{hi}\right)^2.$$

This approximation depends on how well the ratio $\frac{M_h}{M_h'}$ approximates the ratio $\frac{N_h}{n_h}$ for each $h = 1,\ldots,L$. This will depend on how much the M_{hi} vary within each stratum, and by how large n_h is for each stratum. If these ratio approximations are not good, then this formula produces a significantly biased estimate of the mean squared error. We can compare these numbers across all strata by looking at the total squared difference between the two. The square root of this total is the Euclidean distance between the two vectors, so this gives us a way to measure the closeness of the ratios across strata. The table below gives results.

Stratum h	$\frac{M_h}{M'_h}$	$\frac{N_h}{n_h}$	Squared Error
1	8.60	7.25	1.82
2	19.61	18.33	1.63
3	18.04	18.50	0.21
4	24.66	23.50	1.35
5	27.56	27.50	0.00
6	31.75	31.50	0.06
7	42.26	45.50	10.53
8	64.98	64.00	0.96
9	60.94	62.67	2.99
10	106.37	112.33	35.53
11	206.64	186.50	405.50
Total			460.58

Most of the total squared difference comes from stratum 11. So one should question the approximation for the mean squared error of the biased estimator, $v(\hat{Y}_R)$. However, this is not an extremely large total, so the overall approximation may not be all that bad.

To further evaluate the statistical bias, we can use the results of the simulation described in the previous section. For evaluating the bias of $\hat{C}_R$, we compared the average of the 10,000 realizations with the known value of the total in-place cost.

Item	Dollar Value
Total Hardwire In-place Cost[15]	4,781,334,293
Average value of $\hat{C}_R$	4,814,804,896
Standard Error of the Average of $\hat{C}_R$	8,526,987
Estimated Bias	33,470,603
Bias as a Percentage of the Total	0.70%

These results indicate that the bias of the estimator $\hat{C}_R$ may not be that bad. The mean value of the estimator is approximately seven-tenths of a percent above the actual total hardwire investment.

To evaluate the bias in the approximation of the mean squared error of $\hat{C}_R$, we first estimated the mean squared error in the following way:

1. For each of the 10,000 realizations of $\hat{C}_R$, subtract the true total in-place cost from the estimate.

[15] This is the total hardwire investment in the frame used for the simulation.

2. Square each of these errors.
3. Find the average of the 10,000 squared errors.

We then compared the result of this calculation with the average of the 10,000 values of $v(\hat{C}_R)$.

Item	**Value** Squared $ $\times 10^{15}$
Simulation Estimate of the Mean Squared Error	728.1
Standard Error of the Simulation Estimate	10.6
Average value of $v(\hat{C}_R)$	722.8
Standard Error of the Average of $v(\hat{C}_R)$	6.1
Estimated Bias	-5.3
Standard Error of the Estimated Bias	12.4
Bias as a Percentage of the Total	-0.73%

It appears that $v(\hat{C}_R)$ does a fairly good job at approximating the mean squared error. Thus, the biased estimator, $\hat{Y}_R$, in conjunction with its approximate mean square error, $v(\hat{Y}_R)$, are reasonable to use for estimating dollar values from the BellSouth CLLI frame.

Substitution Biases

Aside from statistical calculation issues, there are other sources of bias. We are concerned about statements in the report that refer to the substitution of originally selected items. First, the draft report states on page 7 of Appendix B that

> "In some instances, the location initially selected was impractical to audit, ... In such cases, another location was randomly selected from that stratum."

If the FCC does not want to audit certain locations, their conclusions should be narrowed accordingly – in fact, just to the records in locations that the FCC was willing to audit.

Additionally, footnote 18 in Appendix B talks about substituting for "hard-to-get-to" line items by using the preceding line item in the population list. This changes the probability of selection for certain items, and if this is not accounted for, the result is a bias in the estimate.

Biases Induced by Weaknesses in Audit Controls

It appears that there were procedural weaknesses throughout the audit process. For example, the audit staff did not use the same team of auditors to inspect each location. When examining the proportion of items found by different audit teams, there are noticeable differences in the scoring of line items.

The FCC has twice changed the initial on-site audit scores. These changes were not based on additional information or revisits. They were made back in the office after supervisory review to compensate for inconsistencies in coding. Because of these repeated revisions, there is reason to suspect the accuracy of the entire scoring process.

The FCC may argue that the revisions improve consistency of coding and correct errors. The point is, if there are errors or inconsistencies, how can we be assured all errors and inconsistencies were identified and corrected? Also, were some of the last revisions a result of a policy change in the categorization of certain items previously scored as a "4"? If so, how can we be certain the coding is consistent when the coding criteria has changed long after the onsite FCC visits – many months after, in fact?

The error introduced by incorrect scores is not accounted for in statistical estimates, variance equations, or confidence interval calculations. This is *non-sampling* error because it comes from a source other than the random selection of a sample. It is difficult to quantify non-sampling error. If we compare the estimates based on the current FCC scores, and BellSouth's own scoring of line items in the audit, we see how much of a difference scoring changes can make – approximately $140 million difference in the estimates. Table 5, at the end of this appendix, contains the estimates used for this comparison.

Bayesian Methodology

Here we provide some insight into the general Bayesian approach for analyzing data sampled from a finite population. We give an example for which the distribution of the total cost of the hard-wired COE not found, under a Bayesian model approach, is a sum of independent random variables, one of which is Cauchy distributed. This result reiterates our earlier (frequentist-based) concern over the understated margin of error. Finally, we briefly discuss the claims in the FCC report.

Bayesian Approach

Before setting up the Bayesian model for the total cost, we take a look at the general Bayesian approach for sampling. Let $y = (y_1, \cdots, y_N)$ denote the data of the finite population of interest. Suppose a simple random sampling technique was employed to sample n observations from the finite population. Denote the observed sample of y as y_{obs}, and the unobserved part of y as y_{unobs}. Suppose we are interested in estimating the finite population average, $\bar{y}$, using our knowledge of y_{obs}. The Bayesian approach draw inferences about $\bar{y}$ from the posterior predictive distribution $p(\bar{y}|y_{obs})$. The estimate of $\bar{y}$ is the posterior mean $\hat{\bar{y}} = E(\bar{y}|y_{obs})$ and the standard error of the estimate is the posterior standard error $\sqrt{Var(\bar{y}|y_{obs})}$. Since $\bar{y}$ can be calculated as

$$\bar{y} = \frac{n}{N}\bar{y}_{obs} + \frac{N-n}{N}\bar{y}_{unobs} \tag{1}$$

the distribution that really matters is $p(\bar{y}_{unobs}|y_{obs})$. To obtain the posterior predictive distribution $p(\bar{y}_{unobs}|y_{obs})$, one starts with the specification of the model $p(y_i|\theta)$ and the prior $p(\theta)$. The posterior distribution of θ is

$$p(\theta|y_{obs}) \propto p(\theta)\ p(y_{obs}|\theta) = p(\theta)\prod_{j=1}^{n} p(y_{obs_j}|\theta).$$

Then the posterior predictive distribution $p(\bar{y}_{unobs}|y_{obs})$ is obtained by first conditioning on θ and then averaging over posterior of θ.

$$p(\bar{y}_{unobs}|y_{obs}) = \int p(\bar{y}_{unobs}|\theta)p(\theta|y_{obs})d\theta$$

The following method is sometimes employed to obtain the approximate posterior mean. Simulate θ^{λ}, $l = 1,\cdots,L$, from $p(\theta|y_{obs})$. For each θ^{λ}, draw a vector y_{unobs} from

$$p(y_{unobs}|\theta^{l}) = \prod_{j=1}^{N-n} p(y_{unobs_j}|\theta^{l}),$$

and average over y_{unobsj} to obtain a draw of $\bar{y}_{unobs}$. Using equation (1), we produce a draw of $\bar{y}$. Averaging over the total L draws of $\bar{y}$, we obtain an estimate of $\bar{y}$.

In particular, to illustrate ideas, we could choose $p(y_i|\theta)$ to be a normal distribution

$$y_i|\mu,\sigma^2 \sim N(\mu,\sigma^2)$$

where $(\mu,\sigma^2) = \theta$. With standard noninformative prior, $p(\mu,\sigma^2) \propto \sigma^{-2}$, we have

$$\bar{y}|y_{obs} \sim t_{n-1}(\bar{y}_{obs}, (\frac{1}{n}-\frac{1}{N})s^2_{obs})\ ^{16}. \tag{2}$$

[16] See Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis*, Chapman and Hall, 1998, page 203.

Stratified Random Sampling and Example

In stratified random sampling, the N units are divided into J strata, and a simple random sample of size n_j is drawn using simple random sampling from each stratum $j = 1, \cdots, J$. This design is ignorable given J vectors of indicator variables, $x_1, \cdots, x_J$, with $x_j = (x_{1j}, \cdots, x_{n_j j})$, where $x_{ij} = 1$ when unit i is in stratum j and 0 otherwise. The vector x_j is fully observed as long as we know, for each j, the total number of units N_j in the stratum. The variables x_j incorporate the stratification information.

To estimate the total cost of COE not found, we assume that for each selected central office, all line-items in the office are sampled and the corresponding cost of COE not found are observed. The sample design is simplified to a one stage stratified design. There are total eleven strata. The Bayesian inference for a normal model is as follows. For each strata, let $y_j = (y_{1j}, \cdots, y_{N_j j})$ be the set of population values of the cost of COE not found. Suppose

$$(y_{ij} | x_{ij} = 1, \mu_j, \sigma_j^2) \sim N(\mu_j, \sigma_j^2)$$

Let $\theta_j = (\mu_j, \sigma_j^2)$. Under the assumption that the eleven parameters $(\theta_1, \cdots, \theta_{11})$ have independent distributed priors, we can obtain posterior inferences separately for the parameters in each stratum. Suppose further that for each strata we choose the priors, $p(\mu_j, \sigma_j^2)$, to be the standard noninformative ones, $p(\mu_j, \sigma_j^2) \propto \sigma_j^{-2}$, then using result (2), we have

$$\bar{y}_j | y_{obs_j} \sim t_{n_j - 1}(\bar{y}_{obs_j}, (\frac{1}{n_j} - \frac{1}{N_j}) s_{obs_j}^2), \tag{3}$$

where $\bar{y}_j$ is the average cost in stratum j, y_{obs_j} is the sample in strata j, $\bar{y}_{obs_j}$ is the sample average in strata j and $j = 1, \cdots, 11$. Applying result (3), we have

$$N_j \bar{y}_j | y_{obs_j} \sim t_{n_j - 1}(N_j \bar{y}_{obs_j}, N_j^2 (\frac{1}{n_j} - \frac{1}{N_j}) s_{obs_j}^2). \tag{4}$$

One way to calculate the total cost is

$$C = \sum_{j=1}^{11} N_j \bar{y}_j .$$

This is to say that the total cost of COE not found, C, is just the sum of eleven independent random variables. In the FCC audit sample of BellSouth, six out of the eleven strata have two central offices selected, i.e., $n_j = 2$. For these strata, the posterior predictive distribution

$p(N_j \bar{y}_j | y_{obs_j})$ is a Cauchy distribution because the degrees of freedom is one. The sum of six independent Cauchy random variables is still a Cauchy. The Cauchy distribution is heavy tailed and none of the moments of the distribution exists -- including the mean and variance. Since C is the sum of five independent non-central t random variables and one independent Cauchy, C will not have a mean nor a variance.

If the FCC staff used a Bayesian approach similar to this, and did not realize that the posterior predictive distribution does not have a mean nor variance, then the credibility intervals may not have been properly calculated. Thus, as was the case in the staff's frequentist analysis, the margin of error in the FCC report is understated.

FCC Claims Regarding Bayesian Analysis

The FCC staff make three claims about Bayesian sampling methodology in order to minimize flaws inherent in their original frequentist analysis. We will address each of these assumptions in turn. The staff does not provide the details of the Bayesian structure that was used for their analysis, that is, the assumptions and formulas used to calculate their results. Thus, our comments are based on a common approach employed when Bayesian methods are used in survey sampling.

1. An estimate of the population mean is independent of the choice of sample weights or choice of stratification.

   Bayesian sampling analysis is model oriented. It can, for example, employ a superpopulation from which the finite population – in this case, the CPR database – is a sample. The finite population is physically sampled in order to make inferences.

   Several factors make up a Bayesian sampling model, and one of the key factors is to create a probability distribution to incorporate for prior knowledge about characteristics of the population that is being sampled. The staff has not indicated what distribution they are using for their prior knowledge. If it is a model that assumes a lot of prior knowledge, then justification is needed. A model that assumes little to no prior knowledge generally produces results similar to a frequentist analysis (although interpretation of the results may be different).

   While some Bayesian sampling models may provide estimates which do not rely on the sample design, they are often not independent of the choice of design elements such as stratification.[17] With different strata, the Bayesian estimates would usually be different too.

[17] See Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin, *Bayesian Data Analysis,* Chapman and Hall, 1998, page 224.

In the above example, it is easy to see that the Bayesian estimate of the population mean is not independent of the choice of stratification because y_{obs_j} is dependent on how the strata were formed.

None of this means that the original, frequentist-based analysis was properly done. The sample design is still unsuitable for a precise estimate of the cost associated with unlocatable items, and the margin of error has not been properly calculated. The fact that the staff may have a Bayesian model that provides estimates that are close to improperly calculated values does not provide corroboration.

2. The Bayesian method is design free, so the estimator is unbiased.

   Under frequentist theory, the concept of unbiasedness is that, over repeated sampling, the average of a parameter estimate should be equal to its true value. This is usually considered an intuitively appealing concept, and much theory has been developed concerning minimum variance unbiased estimators.

   In the Bayesian world, one wants to determine the (posterior) distribution of a parameter; the concept of an unbiased point estimate is unimportant. The principle of unbiasedness is reasonable in the limit of large samples, but otherwise is potentially misleading.[18]

   In Bayesian sampling theory terms, minimizing bias will often lead to counterproductive increases in variance. Thus, the FCC staff's statement that since the Bayesian method is design free, the estimator is unbiased seems to us to be irrelevant to the Bayesian analysis that was performed.

   In Bayesian sampling theory terms, minimizing bias will often lead to counterproductive increases in variance. Thus, the FCC staff's statement that since the Bayesian method is design free, the estimator is unbiased seems to us to be irrelevant to the Bayesian analysis that was performed.

   Furthermore, the observed sample y_{obs} along with the likelihood model and prior dictate the Bayesian estimate. If y_{obs} is biased due to the improper sample design, then the Bayesian estimate of the population mean is also biased.

3. The sample mean is the most likely estimate of the population mean.

   In a Bayesian context, the reference to "sample mean" as the most likely value is not clear. We are assuming that the report is referring to the mean of the posterior distribution. While it is true that the Bayesian interpretation of the properties of the estimators is different from a frequentist interpretation, nevertheless, the statement in assumption 2 is not warranted without imposing strong conditions on the prior and the data.

[18] Ibid. Page 108.

If we take this claim to be true, then the posterior predictive distribution has to satisfy one condition: the sample mean coincides with the highest mode of the posterior predictive distribution. For a, the sample mean is usually the posterior mean if the posterior mean exists. In the above example, the noninformative prior that is typical employed was used, but the result is a the posterior predictive distribution with no mean. If a noninformative prior that produces a posterior predictive distribution with a finite mean was used, then that posterior has to be symmetric and unimodal in order for the mean and the mode to coincide. Since FCC report does not provide any information on their posterior predictive distribution, there is no way to check whether the conditions are satisfied to support their claim.

Regardless of what value is "most likely," it is not clear what conclusions, if any, can be drawn from that calculation. In particular, it does not respect the uncertainty in the answer however measured. In our view, the lower bound of a 99% confidence interval remains the best estimate for assessing the total cost of the COE not found.

Calculations

The FCC provided BellSouth with a list of 1831 CLLI along with the record counts for each CLLI. Based on this list and the frame summary table on page 7 of the draft report's Appendix B, we were able to place each CLLI chosen for the audit into the strata defined by the FCC. Table 1 below summarizes this frame, and also shows a comparison with the best match (BM) frame that we used for the simulation described in the "Margin of Error" section.

Calculation Comparisons

Tables 2 and 3 show the results of calculations published in the draft report with our attempt to verify the numbers. Our calculations match almost exactly. Thus, we are confident that our programs use the same equations as the calculation programs the FCC used.

Updated Results

We also present the results (in Tables 4 and 5) of calculating the estimates under four different scenarios:

1. using the biased estimator and its approximate mean squared error with the FCC's current scoring of audited line items (given in the draft report);
2. using an unbiased estimator and its variance with the FCC's current scoring of audited line items (given in the draft report);
3. using the biased estimator and its mean squared error with the BellSouth's scoring of audited line items; and

4. using an unbiased estimator and its variance with the BellSouth's scoring of audited line items.

The unbiased estimator is one that is based on weighting each value in the sample by the inverse of the probability of selection. As with the biased estimator, it can be used to estimate the proportion of compliant line items, or the total in-place cost of non-locatable line items. The formula for the estimator and its variance can be found in Cochran (1977).

One-sided lower (upper) bounds are given for the in-place cost (proportion) estimates. If, for example, the one-sided lower bound of the in-place cost of non-locatable line items was less than 100 million dollars, then one could not statistically conclude, at the appropriate confidence level, that the true value is actually more than 100 million dollars. Similarly, if the one-sided upper confidence bound is greater than 0.90, then one could not statistically conclude, at the appropriate confidence level, that the true value is actually less than 0.90 (90 percent).

To calculate one-sided lower confidence bounds for all proportion estimates and the unbiased in-place cost estimates, Student's t distribution was used with the effective degrees of freedom calculated using the Satterthwaite approximation. For the biased estimates of in-place cost, the multiplying factors determined by the simulation results in the "Margin of Error" section were used.

In general, the one-sided lower 99 percent bounds for the in-place cost of non-locatable items are either below zero, or are a minute fraction (about four-tenths of a percent)of the BellSouth's total hardwire investment.[19] At the 95 percent confidence level, the one-sided lower bounds are still only a small fraction of the total hardwire investment, especially when BellSouth's scores of audited line items are used (about eight-tenths of a percent).

[19] This assumes that total hardwire investment is $8.778 billion, as stated in the table on page 4 of the draft audit report.

Table 1 - FCC vs. "Best Match" (BM) Sampling Frame Summary

Stratum h	Number of Records per Central Office Location in Stratum Both Frames		Number of Central Office Locations in Stratum N_h		Number of Central Office Locations Selected for Audit n_h		Number of Records per Stratum M_h	
	High	Low	FCC	BM	FCC	BM	FCC	BM
1	5185	2005	29	334	4	4	86772	1060527
2	1994	1014	110	232	6	6	149554	337802
3	993	900	37	22	2	2	35042	20658
4	899	801	47	33	2	2	40150	28098
5	797	705	55	43	2	2	41286	32135
6	699	610	63	57	2	2	41407	37203
7	598	500	91	94	2	2	49904	51859
8	493	401	128	119	2	2	56336	52719
9	399	300	188	188	3	3	65569	65796
10	299	200	337	199	3	3	83396	50236
11	199	100	746	154	4	4	104765	24098
Total			1831	1475	32	32	754181	1761131

Table 2 - Calculation Verification
Percentage of Compliant Records

Item	Estimate	Standard Error	Margin of Error[20]	Lower Confidence Bound[20]	Upper Confidence Bound[20]
FCC Published	80.50	1.53	2.99	77.51	83.49
Verification Results	80.50	1.51	2.97	77.53	83.46

Table 3 - Calculation Verification
Total In-Place Cost ($M) of Non-Locatable Line items

Item	Estimate	Standard Error	Margin of Error[20]	Lower Confidence Bound[20]	Upper Confidence Bound[20]
FCC Published	291.7	72.9	142.9	148.8	434.6
Verification Results	291.5	72.9	142.9	148.7	434.4

[20] The margin of error was determined using the 97.5th percentile of a standard normal distribution. This is the same methodology used in the draft report. While we do not feel that this is the correct thing to do, it was done here in order to make a comparison. The resulting error bounds give a 95 percent confidence interval.

Table 4 - Percent of Compliant Records

Scoring System	Estimator Type	Estimate	Standard Error	Effective Degrees of Freedom	One-Sided 95% Upper Bound	One-Sided 99% Upper Bound
	Biased	80.50	1.51	11.67	83.22	84.61
FCC	Unbiased	77.70	2.11	12.15	81.45	83.34
	Biased	85.16	1.08	9.32	87.14	88.21
BellSouth	Unbiased	82.19	2.21	10.45	86.19	88.29

Table 5 - Total In-Place Cost ($M) of Non-Locatable Line items

Scoring System	Estimator Type	Estimate	Standard Error	Effective Degrees of Freedom	One-Sided 95% Lower Bound	One-Sided 99% Lower Bound
	Biased	291.5	72.9	N/A	114.1	-32.6
FCC	Unbiased	284.6	82.5	7.77	128.4	37.4
	Biased	147.9	33.7	N/A	65.9	-2.0
BellSouth	Unbiased	151.7	41.8	7.99	72.4	26.3

4

Exhibit 4

FEDERAL COMMUNICATIONS COMMISSION
WASHINGTON, D.C. 20554

IN REPLY REFER TO:
1600E3

March 11, 1998

Ms. Mary L. Henze
Director- Executive and Federal Regulatory Affairs
BellSouth Corporation
1133 21 Street, N.W. Suite 900
Washington, D.C. 20005

Dear Ms. Henze:

We have tabulated the results of our recent field work involving the audit of BellSouth Telecommunications' (BST) Continuing Property Records. We have analyzed the evidence collected by the auditors and have looked at this in conjunction with an examination of the full CPR for each of the locations. Based upon this analysis, and with the application of consistent decision-making rules across all sampled units, we have evaluated each of the 36 line items inventoried within the 32 audited sites. In some cases our preliminary field assessments have changed as the result of our subsequent analyses.

The current assessments for the 32 offices are summarized on the attached pages. In order for us to consider any changes to these findings, we will need the necessary documentation that you believe will support your position by April 7, 1998. Also, based on a recent conversation with Howard Burnette, we expect any documentation that you currently have that supports BST's position to be sent to us immediately. If you have any questions, please contact Wesley Jarmon at (202) 418-0815.

Sincerely,

Kenneth M. Ackerman

Kenneth M. Ackerman
Chief, Audits Branch

Attachments

5

PRICEWATERHOUSECOOPERS

September 22, 1999

PricewaterhouseCoopers LLP
1100 Campanile Building
1155 Peachtree Street
Atlanta GA 30309-3630
Telephone (404) 870 1100
Facsimile (404) 870 1239

Mr. Isaiah K. Harris, Jr.
Vice President and Chief Financial Officer
BellSouth Telecommunications, Inc.
4503 BellSouth Center
675 West Peachtree Street, NE
Atlanta, Georgia 30375

Dear Mr. Harris:

At BellSouth management's request, PricewaterhouseCoopers LLP performed procedures to assess the results of the BellSouth Telecommunications Inc.'s ("BST" or "the Company") Continuing Property Records ("CPR") Audit conducted by the Common Carrier Bureau ("Bureau"), which was released on March 12, 1999.

As background, the Bureau conducted an audit related to hard-wired central office equipment ("COE"). The purpose of the audit, as stated by the Bureau, was to determine whether BST's plant records: a) were in compliance with the Federal Communications Commission's requirements regarding basic property records and CPR, as set forth in sections 32.2000 (e) and (f) of the Commission's rules, and b) to determine whether BST's plant accounts accurately reflected the cost of assets used and useful in the provision of telecommunications services.

In order to test compliance with these rules, the Bureau selected 1,152 items (36 items from 32 central offices) from a population of 754,181 items contained in the BellSouth CPR as of July 31, 1997. The Bureau performed its audit procedures on each of the sampled items and utilized a rating scheme in which sampled items were placed into one of four categories (1-4). The following describes what each category represented:

1. Found
2. Found in another location
3. Not found
4. Unverifiable

The Bureau provided BST with its results and requested BST to correct and comment on factual errors and omissions in the audit. The Bureau made some adjustments to the results based upon data provided by BST and then released a final report, in which the Bureau summarized their findings in the following categories:

- Asset Found (AF) – encompassed all items finally scored as 1.
- Asset Partially Found (APF) – encompassed 20 of the 116 items finally scored as 3.
- No Asset Found (NAF) – encompassed 96 of the 116 items finally scored as 3.
- Unverifiable Asset (UA) – encompassed all the items scored as 2 or 4.